Optimal State Estimation in High Noise

B. D. O. ANDERSON AND J. B. MOORE

*Department of Electrical Engineering, University of Newcastle,
New South Wales, Australia*

The problem is examined of estimating the state of a linear dy-
namical system in the presence of high measurement noise. It is con-
cluded that optimal filter design may be simplified to the extent that
it need not depend on the solution of a matrix Riccati differential
equation, but only on the solution of a matrix linear differential equa-
tion. For a related problem, that of estimating a signal $s(t)$ given
noisy measurements $s(t) + w(t)$ where the noise is large and the co-
variance of $s(t)$ is known, optimal filter design is immediate.

INTRODUCTION AND BACKGROUND

The problem of detecting faint messages in a high noise level was raised
by Wiener (1949). More specifically, a function of time $z(\cdot)$ is given
which is a measurement of a signal $s(\cdot)$ corrupted by noise $w(\cdot)$; thus
$z(t) = s(t) + w(t)$. The noise is assumed to have a spectral density
$\Phi_{ww}(j\omega) = 1$ and the signal to have a spectral density $\Phi_{ss}(j\omega) = \epsilon F(\omega)$
where ϵ is a small number. A procedure is given for generating an optimal
(minimum-variance, unbiased) estimate of $s(t)$ from the measurement
$z(\tau)$, for $-\infty \leq \tau \leq t$.

Subject to certain constraints it turns out that $s(\cdot)$ may be generated
by subjecting $z(\cdot)$ to a delay followed by passage through a filter with a
transfer function $\epsilon F(\omega)$, or for more accuracy, through a filter with trans-
fer function $\epsilon F(\omega)[1 + \epsilon F(\omega)]^{-1}$.

The design procedure for the optimal filter is evidently far simpler as
a result of the high noise assumption than the usual design procedure
encountered.

In this paper we are concerned with expanding Wiener's results to cope
with time-varying systems, nonstationary stochastic processes, and an
initial time which need not be $-\infty$. The principle conclusion drawn is
that the design of an optimal filter need not depend on the solving of a
Riccati differential equation (Kalman and Bucy, 1961), but rather a

linear differential equation. This represents a substantial potential saving in the computation time for the design procedure.

In Section 2, a problem formulation is given, while Section 3 deduces the design of the simplified optimal filter. Section 4 shows how when the covariance of $s(t)$ is specified, estimation of $s(t)$ need not even require the solution of a linear differential equation.

One feature of the new results is the elimination of the requirement to use a delay in the approximately optimum filter. Actually, Wiener (1945) suggests elsewhere than in the treatment of high noise level filtering that filtering problems may be approached either by allowing a delay, or by performing a spectral factorization. The material presented herein in effect corresponds to taking the latter course, i.e., performing a spectral factorization, insofar as this term may be applied to time-varying problems.

2. PROBLEM FORMULATION

The situation depicted in Fig. 1 will be considered. The variables v and w denote independent white noise with cov $[v(t), v(\tau)] = Q(t)\delta(t - \tau)$ and cov $[w(t), w(\tau)] = R(t)\delta(t - \tau)$ [where $R(t)$ is a positive definite matrix for all t]. The full equations of the system depicted are

$$\dot{x} = Fx + Gv; \qquad z = H'x + w. \tag{1}$$

The standard problem is to estimate $x(t)$ from a knowledge of $z(\tau)$ for $t_0 \leq \tau \leq t$ and any data concerning $x(t_0)$, which will be assumed here to be a random variable of zero mean and known covariance. The estimate of $x(t)$, written $\hat{x}(t)$, is required to be linear, of minimum variance, and to be unbiased. The estimate $\hat{x}(t)$ may be found from the linear system of
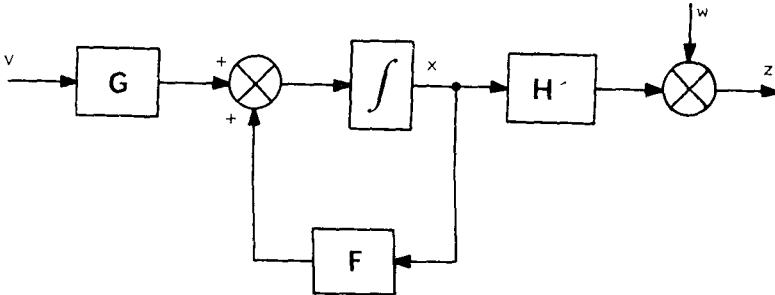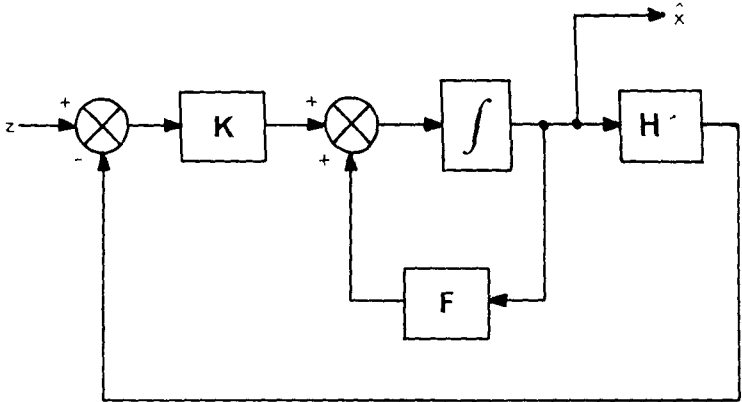

Fig. 1. Prescribed system.

FIG. 2. Kalman filter.

Fig. 2 (Kalman and Bucy, 1961) where the gain matrix $K(t)$ is determined as follows:

$$K(t) = P(t)H(t)R^{-1}(t), \tag{2}$$

where $P(t)$ satisfies the nonlinear equations

$$\dot{P} = PF' + FP - PHR^{-1}H'P + GQG' \tag{3}$$

with boundary condition $P(t_0) = \text{cov}\,[x(t_0), x(t_0)]$. If either F is uniformly asymptotically stable or $[F, H]$ is completely observable, $t_0 = -\infty$ is allowed; if $[F, H]$ is uniformly completely observable, $[F, G]$ is uniformly completely controllable, and F, Q and R are bounded, then the optimal filter is uniformly asymptotically stable.

The evaluation of solutions of (3) clearly presents some computational difficulties; we shall show how many of the difficulties are eliminated in the high noise case.

With $y^* = H'x$, the high noise case becomes one where the covariance of w is always much greater than that of y, in the sense that for all t_1, t_2 within the filtering time interval and all $u(\cdot)$,

$$\int_{t_1}^{t_2} \int_{t_1}^{t_2} u'(t)E[w(t)w'(\tau)]u(\tau)\,dt d\tau$$

$$\geqq \int_{t_1}^{t_2} \int_{t_1}^{t_2} u'(t)E[y(t)y'(\tau)]u(\tau)\,dt d\tau. \tag{4}$$

Now it is readily checked that

$$E[y(t)y'(\tau)] = H'(t)\Phi(t, \tau)P_s(\tau)H(\tau)1(t - \tau)$$
$$+ H'(t)P_s(t)\Phi'(\tau, t)H(\tau)1(\tau - t) \qquad (5)$$

where $\Phi(\cdot, \cdot)$ is the transition matrix of $\dot{x} = Fx$, and $P_s(\cdot)$ is the state covariance matrix, given by

$$\dot{P}_s = P_s F' + FP_s + GQG', \qquad (6)$$

We distinguish the case of filtering over a finite interval $[t_0, T]$ and over a semi-infinite or infinite interval. In the former case, suppose F, G, H, and Q are bounded. Then one can readily find a number η such that

$$\| E[y(t)y'(\tau)]\| \leq \eta$$

for all t, $\tau \in [t_0, T]$ where $\|A(t)\|$ denotes a pointwise norm, e.g., $[\text{trace } A'A]^{1/2}$. The high noise constraint (4) becomes

$$\lambda_{\min}[R(t)] \gg \eta \qquad (7)$$

where $\lambda_{\min}$ is the minimum value for $t \in [t_0, T]$ of the minimum eigenvalue of $R(t)$.

For semi-infinite or infinite interval filtering, we require F, G, H, and Q to be bounded, and F to be exponentially asymptotically stable. Then one can readily find positive constants C_1 and C_2 so that

$$\|H'(t)\Phi(t, \tau)P_s(\tau)H(\tau)\| \leq C_1 e^{-C_2(t-\tau)}. \qquad (8)$$

Now

$$\left| \int_{t_1}^{t_2} \int_{t_1}^{t_2} u'(t)E[y(t)y'(\tau)]u(\tau) \, dt d\tau \right|$$

$$= 2 \left| \int_{t_1}^{t_2} u'(t) \left[\int_{t_1}^{t_2} H'(t)\Phi(t, \tau)P_s(\tau)H(\tau)u(\tau) \, d\tau \right] dt \right|$$

$$\leq 2 \int_{t_1}^{t_2} \|u'(t)\| \frac{C_1}{(2C_2)^{1/2}} \left[\int_{t_1}^{t_2} \|u(\tau)\|^2 \, d\tau \right]^{1/2} dt$$

$$\leq C_1 \left(\frac{2}{C_2}\right)^{1/2} \int_{t_1}^{t_2} \|u(\tau)\|^2 \, d\tau$$

(where the second step above follows as in [Anderson, 1968]). Thus the

high noise constraint (4) becomes

$$\lambda_{\min}[R(t)] \gg C_1(2/C_2)^{1/2}. \tag{9}$$

Notice that in the absence of the constraints on F, G, H and Q the covariance of $y(t)$ may not be bounded over the filtering interval; the constraints imposed are reasonable on physical grounds.

3. APPROXIMATE STATE ESTIMATION

A heuristic approach to the development of a state estimator will first be given which is based on the solution of the Wiener–Hopf equation. The impulse response $A(t, \tau)$ of an optimal estimator is given by the integral equation, valid for $t > \tau \geqq t_0$

$$\text{cov}\,[x(t), z(\tau)] = \int_{t_0}^{t} A(t, \sigma)\,\text{cov}\,[z(\sigma), z(\tau)]\,d\sigma. \tag{10}$$

We make the heuristic assumption that in a high noise situation, $\text{cov}\,[z(\sigma), z(\tau)]$ in the above equation may be replaced by $\text{cov}\,[w(\sigma), w(\tau)]$, i.e. $R(\sigma)\delta(\sigma - \tau)$. With this replacement (10) yields $\text{cov}\,[x(t), z(\tau)] = A(t, \tau)R(\tau)$. Now

$$\begin{aligned}
E[x(t)z'(\tau)] &= E[x(t)x'(\tau)H(\tau)] + E[x(t)w(\tau)] \\
&= \Phi(t, \tau)P_s(\tau)H(\tau)
\end{aligned} \tag{11}$$

as may easily be shown. Thus

$$A(t, \tau) = \Phi(t, \tau)P_s(\tau)H(\tau)R^{-1}(\tau). \tag{12}$$

This suggests that a suitable state estimator is provided by the arrangement of Fig. 3. Although the above argument is not rigorous, a more rigorous approach to follow will confirm the plausibility of (12). There is also a parallel between (12) and the Wiener result, in the sense that (12) indicates that the optimum filter is almost derivable from the antecedal part of the covariance (5). This covariance corresponds to $\epsilon F(\omega)$ in the Wiener theory; whereas the Wiener theory does not select the antecedal part of the inverse Fourier transform of $\epsilon F(\omega)$, the use of a delay element is a specific technique for avoiding the extraction of the antecedal part.

The importance of (12) is that the calculation of the P matrix required for the scheme of Fig. 2 via the *nonlinear* differential equation (3) is replaced by the calculation of the P_s matrix, using the *linear* differential equation (6).
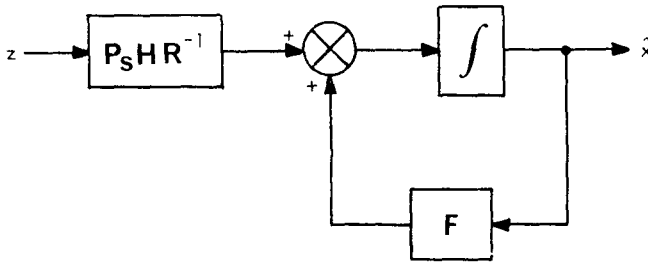
Fig. 3. Approximate optimum filter derived from heuristic argument.

Let us now proceed with a more accurate justification of the results; at the same time the approximate filter will turn out to be slightly different to that of Fig. 3, in fact paralleling the Wiener theory transfer function $\epsilon F(\omega)[1 + \epsilon F(\omega)]^{-1}$, rather than $\epsilon F(\omega)$.

It is pointed out by Kalman and Bucy (1961) and by Kalman (1961) that P in (3) is a nonnegative-definite matrix, and that P_s of (6) is an upper bound for P in the sense that $P_s - P$ is nonnegative-definite. Accordingly, because P_s is bounded independently of R, P is bounded, irrespective of what R is.

Now define

$$W = PHR^{-1}H'P \qquad (13)$$

and form, from (3) and (5),

$$(\dot{P}_s - \dot{P}) = F(P_s - P) + (P_s - P)F' + W, \qquad (14)$$

with $P_s - P \mid_{t_0} = 0$. The solution of (14), regarded as an equation for the unknown $P_s - P$, with known F and W, is easily found as

$$P_s(t) - P(t) = \int_{t_0}^{t} \Phi(t, \lambda) W(\lambda) \Phi'(t, \lambda) \, d\lambda. \qquad (15)$$

It follows that

$$\| P_s(t) - P(t) \| \leq \max_{\lambda \in [t_0, t]} \| W(\lambda) \| \int_{t_0}^{t} \Phi(t, \lambda) \Phi'(t, \lambda) \, d\lambda. \qquad (16)$$

Because F is bounded, and exponentially asymptotically stable in semi-infinite interval problems, the integral in (16) is bounded independently of t, and whether or not $t_0 = -\infty$. Also, because P and H are bounded,

irrespective of R, it follows that

$$\| P_s(t) - P(t) \| \leqq M \{ \lambda_{min}[R(t)] \}^{-1} \tag{17}$$

for some positive constant M, determinable from the known bounds on F, G, H and Q and the exponential bound on $\Phi(\cdot , \cdot)$.

Consequently, for the high noise case, $P_s(t) - P(t) \cong 0$ or $P_s(t) \cong P(t)$ and $P_s(t)H(t)R^{-1}(t) \cong P(t)H(t)R^{-1}(t)$. This equation moreover holds uniformly in t, i.e. for sufficiently high noise, the approximation may be made arbitrarily good; the equation also suggests that the optimum estimator of Fig. 2 may be replaced by the approximate estimator of Fig. 4.

To see that close approximation of the gain $K(t)$ means close approximation of $x(t)$, observe first that, for a finite interval $[t_0, T]$, the result is trivial; for a nonfinite interval, it follows given uniform asymptotic stability of the optimal filter and the boundedness of all the filter parameters (conditions guaranteeing these filter properties appear earlier). Under such constraints, the filter is structurally stable, i.e., small variations in the filter elements preserve approximately the desired behavior of the filter from the input–output point of view.

For the same sort of reason, we can see why the unity negative feedback loop in Fig. 4 might be eliminatable, to yield Fig. 3 again.

This loop substracts a signal $H'(t)\hat{x}(t)$ from $z(t)$. Now $H'(t)\hat{x}(t)$ is roughly like $H'(t)x(t)$, which is known to be much "smaller" than $z(t)$. Thus it could be argued that the feedback of this signal could be dis-
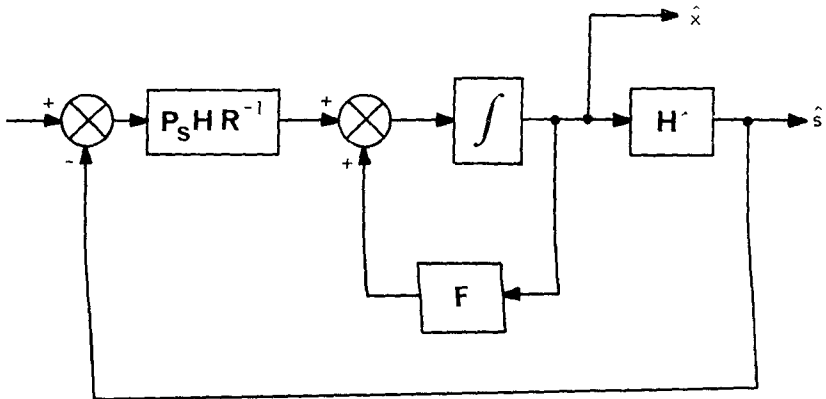

FIG. 4. Approximate optimum filter derived from exact arguments.

pensed with. Note the fallacy in this argument though: it is the small part of $z(t)$, viz. $H'(t)x(t)$, as distinct from the "large" part, viz. $w(t)$, which makes any filtering possible.

4. ESTIMATION GIVEN THE SIGNAL COVARIANCE

To achieve a slightly closer parallel with Wiener's result, and to show how under some condition filter calculation can be simplified even further than has been hitherto considered, we can pose and solve the following related problem.

Suppose noisy measurements $z(t)$ of a signal $s(t)$ are available, with $z(t) = s(t) + w(t)$ where $w(t)$ is as before, and the covariance of $s(t)$ is given as

$$R_{ss}(t, \tau) = H'(t)\Phi(t, \tau)L(\tau)1(t - \tau)$$
$$+ L'(t)\Phi'(\tau, t)H(\tau)1(\tau - t), \qquad (18)$$

where Φ is the transition matrix of the exponentially asymptotically stable system $\dot{x} = Fx$, and F, H and L are bounded with appropriate controllability and observability conditions holding. How, from z, might $s(t)$ be best estimated?

Suppose, for the moment, that $s(t)$ is the output of some linear dynamical system excited by white noise. The "G matrix" of such a system is hard to determine, but the form of $R_{ss}(t, \tau)$ determines the "F" and "H matrices" immediately, as $[(d/dt)\Phi(t, \tau)]\Phi^{-1}(t, \tau)$ and $H(t)$, respectively. To estimate $s(t)$, which is a variable obtained by transforming the system states by $H'(t)$, one simply also transforms an estimate of the system states by $H'(t)$; thus the filter of Fig. 4, with $\hat{s}$ as the output, constitutes the estimate of s.

The important point to observe is that design of this filter is immediate. This is because the gain P_sHR^{-1} is defined immediately on comparing (5) and (18) as LR^{-1}. In other words, the construction of a system generating $s(t)$ is not required, either physically or in order to compute P_s, because the quantity of interest, viz. P_sH, is already known from specification of R_{ss}.

Thus to design the optimal filter, not even a linear differential equation has to be solved. Note also that without the unity negative feedback, the optimum filter is simply a cascade of a scaling element R^{-1} and a filter with impulse response equal to the antecedal portion of $R_{ss}(t, \tau)$.

The extent to which an arbitrary $R_{ss}(t, \tau)$ can be written in the form

of (18) with all desired conditions holding on F, L, etc., may be gleaned from Silverman (unpublished) and Silverman and Anderson (1968). The latter can be decomposed as $H'(t)\Phi(t, \tau)L(\tau)$, with $\Phi(\cdot, \cdot)$ a transition matrix of $\dot{x} = Fx$, all of F, L and H bounded, $[F, L]$ uniformly completely controllable and $[F, H]$ uniformly completely observable. Silverman and Anderson (1968) show that if $A'(t)B(\tau)1(t - \tau)$ is a bounded-input, bounded-output impulse response matrix, then the procedures of Silverman (unpublished) lend to an asymptotically stable F matrix.

Similar simplification will be observed for high noise filtering in discrete time systems; thus the recursive formulas used for computing the optimal filter gain will drastically simplify.

REFERENCES

WIENER, N. (1949). "Extrapolation, Interpolation and Smoothing of Stationary Time Series." M.I.T. Press, Cambridge, Massachusetts.

KALMAN, R. E. AND BUCY, R. S. (1961). New results in linear filtering and prediction theory. *J. Basic Engineering Trans. Am. Soc. Mech. Eng., Ser. D* **83,** 95–108.

ANDERSON, B. D. O. Input-output properties of certain stable systems. *Proc. IEEE* **56,** 869–870.

KALMAN, R. E. (1961). New methods and results in linear filtering and prediction theory, *in* "Proceedings of the Symposium on Engineering Applications of Probability and Random Functions." Wiley, New York.

SILVERMAN, L. M. (1968). Synthesis of Impulse Response Matrices by Internally Stable and Passive Realizations (unpublished).

SILVERMAN, L. M. AND ANDERSON, B. D. O. Controllability, observability and stability of linear systems, *SIAM J. Control* **6,** 121–130.